The Semantic Network: Uncovering the Mechanisms of New Language Integration

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in the undergraduate colleges of The Ohio State University

Apoorva Vallampati

The Ohio State University

December 2020

Advisors:

Dr. Layla Unger, The Cognitive Development Lab, Department of Psychology

Dr. Olivera Savic, The Cognitive Development Lab, Department of Psychology

Dr. Vladimir Sloutsky, The Cognitive Development Lab, Department of Psychology

Abstract

Semantic relations between words (e.g., between *drink* and *soda*) are crucial for language fluency. Language is replete with statistical regularities from which people can potentially form these links. We focus on two such regularities: direct co-occurrence and shared co-occurrence. Words that appear together in sentences and express meaningful ideas (e.g., *drink-soda*) tend to reliably *directly co-occur* together, and words similar in meaning tend to *share* patterns of direct co-occurrence across sentences (e.g. *soda* and *milk* share co-occurrence with *drink*). In this study, we investigate which of these regularities children (4-year-olds) and adults can capitalize on to form new semantic links between novel and familiar words. Participants heard sentences in which novel words either directly co-occurred or share co-occurrence with familiar words in a training phase. We then assessed the formation of direct and shared semantic links using an explicit labeling measure. Results suggest that children are sensitive only to direct co-occurrence regularities to form new semantic links, while adults are sensitive to both direct and shared co-occurrence regularities when forming new semantic links. This research is therefore uncovering the development of the mechanisms of semantic organization from mere exposure to language.

Keywords: lexico-semantic development, co-occurrence regularities, novel word learning

The Semantic Network: Uncovering the Mechanisms of New Language Integration

We begin life as newborn infants with no knowledge of language, and grow into adults with the capacity to use language to communicate and understand a virtually limitless variety of meaningful ideas. An important part of developing the ability to express and understand ideas is forming networks of words connected by meaningful links. **For example, meaningful links between *drink* and *soda* can allow us to express and understand ideas about drinking soda.** Impressively, new words are often semantically integrated just by encountering them in language, without the need for explicit teaching. What are the underlying mechanisms with which we incorporate new words into our existing semantic network throughout development?

This study will investigate the development of the ability to build semantic networks from one potentially vital source of input: regularities in language with which words co-occur. In language, words occur in sentences and across sentences in consistent patterns with one another. These regularities may be capitalized on to add new words into our pre-existing semantic networks. We focus on two co-occurrence regularities that can support links between words: **direct co-occurrence** between words that occur together in language (e.g., *drink* and *soda*), and **shared co-occurrence** between words that share each other's co-occurrence with other words (e.g., between *soda* and *milk*, which share co-occurrence with *drink*). See Figure 1 for details.

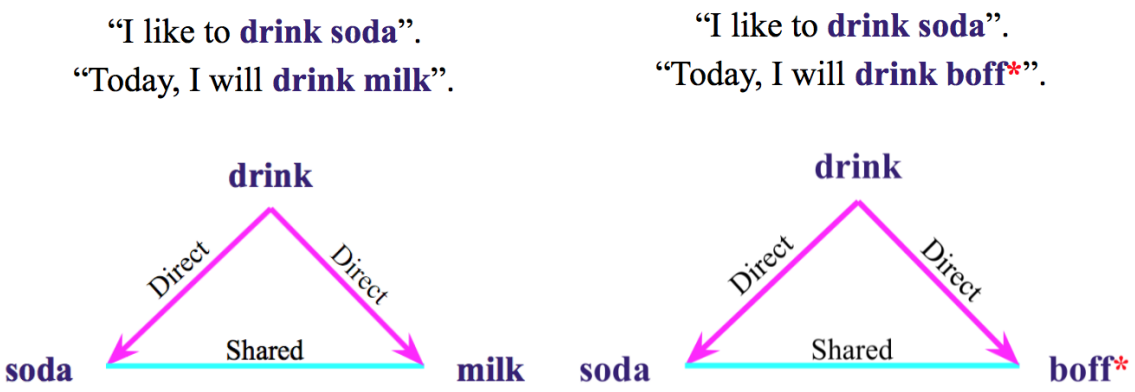

Figure 1. On the left, a pair of sentences containing both direct and shared co-occurrence regularities between familiar words. On the right, the same sentences showing these regularities between familiar and novel words.

We know, from the influential work of Landauer and Dumais (1997) and others (e.g., Jones and Mewhort, 2007), that computer-simulated mechanisms that use these co-occurrence regularities to form links between words build human-like semantic networks. More recent evidence shows that these two types of statistical regularities in language can be capitalized on to build new words into the lexico-semantic networks of adults (Savic and Unger, 2019). Interestingly, co-occurrence links may potentially make important contributions to learning new words, even without other information about what new words mean. For example, in Figure 1, we read the sentence "I like to *drink soda*", in which *drink* and *soda* directly co-occur. When presented with the sentence "Today, I will *drink boff*", readers can make inferences about the meaning of the new word *boff* based on its direct co-occurrence link to *drink*, as well as its shared co-occurrence link to *soda* (*soda* and *boff* both directly co-occur with *drink* in different sentences). Because of these links, we may inherit some knowledge of *boff* as a type of beverage (like *soda*) or, more generally, as a thing in the same lexical category as *soda* (a noun).

Are both children and adults equally sensitive to these two types of co-occurrence regularities? Moreover, can they use co-occurrence regularities with familiar words to integrate new words into their lexico-semantic networks? The present research starts with these critical questions.

Direct Co-occurrence Regularities

The power of direct co-occurrence regularities on learning in domains other than language has been previously studied. This research will lay the foundation on which our research hypothesis is based. These studies provide evidence that the ability to form links based on direct co-occurrence may begin to emerge early in development. One study investigating the development of the ability to use direct co-occurrence to build links showed that infants link sounds that directly co-occur after only 2 minutes of exposure (Saffran, Aslin, & Newport, 1996), and that infants link visual images that directly co-occur in pairs at a high rate (Fiser & Aslin, 2002). A more recent study found that toddlers treat novel words that directly co-occur in sentences as words that are linked (Wojcik & Saffran, 2015). In domains like perception and the processing of objects, space, and time, studies find that toddlers show a sensitivity to predictable, statistically regular events that may support word learning (Saffran and Benitez, 2018). Based on these discoveries, we can formulate a hypothesis that 4-year-olds may be sensitive to direct co-occurrence regularities in language that guide novel word learning.

Shared Co-occurrence Regularities

In contrast to direct co-occurrence sensitivity, a consideration of evidence for shared co-occurrence sensitivity reveals that this ability may develop gradually over time, and may not be present at the beginning of a child's development. We begin our exploration of previous research with a study by Bauer and San Souci (2010) which focused on two groups, 4-year-olds and 6-

year-olds, and examined age-related differences in their abilities to extend their knowledge by integrating new information presented in separate instances. The mode of study here was a read-aloud activity in which one new fact was embedded in each of two passages. Researchers then tested each age group's level of integration of the facts by asking 6-year-olds to answer open-ended questions and 4-year-olds to identify the correct answers without having to verbally answer. The results showed that 6-year-olds were better at integrating facts across separate instances than 4-year-olds were. In essence, information learned across episodes became linked in memory, and the older children were better able to capitalize upon those links to extend their knowledge (Bauer & San Souci, 2010). Further studies show that by age 8, this ability was strengthened even more (Bauer & Larkina, 2017). By broadening the age range into adulthood, researchers observed performance gains on cross-episodal tasks all the way up to the age of 30 (Schlichting, Guarino, Schapiro, Turk-Browne, & Preston, 2017). Using these studies as support, we see that the ability to form meaningful links across different episodes appears to develop with maturation. Thus, we can continue to construct our hypothesis to take into account the general developmental pattern of the ability to form links across episodes, which is necessary to form links between words that share co-occurrence in language.

So far, we've learned that a two-year age difference between 4-year-olds and 6-year-olds accounts for a significantly higher level of spontaneous integration of information across different episodes. We've postulated that shared co-occurrence sensitivity, which requires links between words to be formed across different episodes of exposure to language, will also develop with age. Is shared co-occurrence sensitivity strong enough by adulthood to foster the formation of new links between words? Previous evidence suggests that by the time people are adults, they are able to form links between new words based on their shared co-occurrence across sentences.

A study conducted by McNeill (1963) consisted of a training phase where participants were presented with new words within English sentence frames, in which a pair of new words shared each other's pattern of direct co-occurrence with another new word across sentences. McNeill then administered a free association task, a technique that measures whether two words are linked, to test if participants responded with one word when prompted with the other. The free association responses suggested that adults can form links between words based on both direct and shared co-occurrence regularities in language (McNeill, 1963; McNeill 1966).

In review, Bauer and San Souci (2010) found that the general developmental pattern of integration of shared facts across episodes improved with age. We can apply this finding to shared co-occurrence sensitivity in 4-year-olds and theorize that the development of links based on shared co-occurrence regularities across language will be quite difficult for them, as their sensitivity to this type of regularity is likely poor and will improve with time. The results of McNeill's study suggest that by adulthood, people are able to form links based on shared co-occurrence, further strengthening our theory of shared co-occurrence sensitivity as a cognitive skill that develops over time.

Current Research

Previous studies indicate that young children may be sensitive to direct co-occurrence regularities in language. They also show that adults are sensitive to both direct and shared co-occurrence regularities in language (McNeill, 1963; Sloutsky, Yim, Yao, & Dennis, 2017; Savic, Unger, & Sloutsky, 2020). Sensitivities to statistical regularities in language may also be a powerful force behind novel word learning in that these regularities might serve as a starting point in our semantic understanding of new words due to their relationships with other words, either within or across sentences. However, are children and adults equally sensitive to forming

links based on direct and shared co-occurrence regularities in language? How do these two types of regularities inform the meaning of novel words in children and adults? We investigated these questions by measuring whether children and adults link novel words to familiar words after hearing sentences in which these words either directly co-occur or share co-occurrence. Our prediction is that children and adults are not equally sensitive to picking up on the statistical regularities presented in language. Prior evidence suggests that adults can utilize both direct and shared co-occurrence, while children may only be able to use direct co-occurrence to learn new words.

This study investigated the research question by crafting linguistic input rich in direct and shared co-occurrence regularities using words that were familiar to children (*apple* and *horse*), as well as novel words with which children were unfamiliar (*foobly*, *dodish*, *mipp*, and *geck*). We first exposed 4-year-olds and adult participants to this input through sentences that they heard while watching a video. We then tested whether they learned the novel words based on their direct or shared co-occurrence with the familiar words *apple* and *horse*.

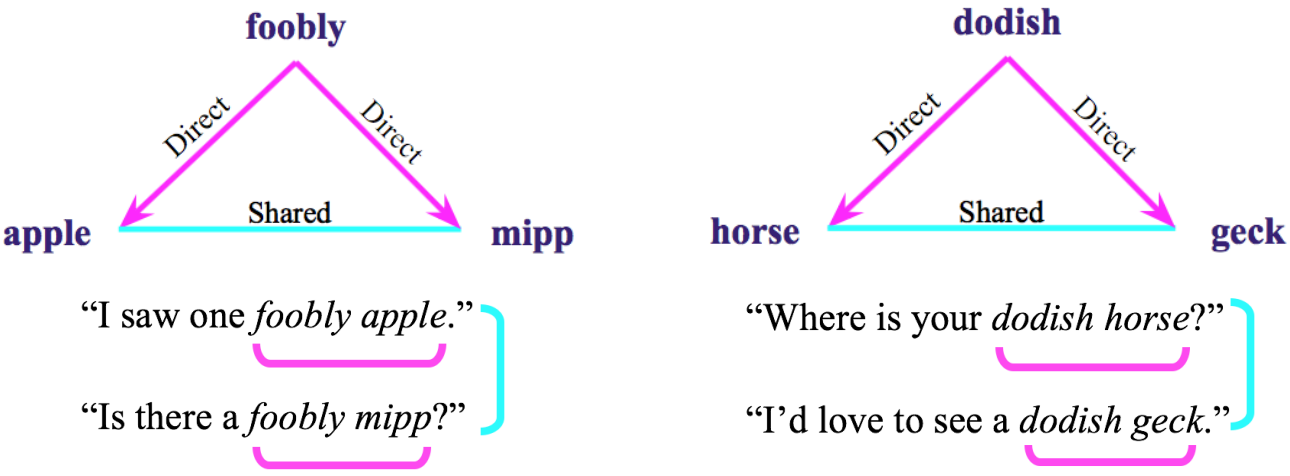

Figure 2. Two depictions of the training triads, each containing two direct co-occurrence regularities and one shared co-occurrence regularity. Beneath them, illustrations showing the ways in which these regularities were inserted into sentences.

This study was divided into two phases: a *training phase* and a *label extension task*. During the *training phase*, participants heard a recording of linguistic input in which the familiar words *apple* and *horse* each directly co-occurred with one novel word (*foobly*; *dodish*), and shared patterns of co-occurrence with another novel word (*mipp*; *geck*). See Figure 2 for details. During the *label extension task*, participants were shown two sets of images: images of *apples* and *horses*, as well as images of items from the same semantic category as apples and horses (fruits and large mammals), like grapes and cows. Participants were asked if an image depicting a familiar word (e.g. apple) was *foobly* or *dodish* to test integration of novel with familiar words via direct co-occurrence, and also if the image was *mipp* or *geck* to test integration via shared co-occurrence. If a link had formed between the familiar words (*apple* or *horse*) and the novel words that directly co-occurred (*foobly* or *dodish*), the participant would choose the correct novel word to label the image. In other words, they would correctly choose *foobly* when they saw an apple and *dodish* when they saw a horse. If a link had formed between the familiar words (*apple* or *horse*) and the novel words that shared co-occurrence with (*mipp* or *geck*), then the participant would choose the correct novel word to label the picture. In other words, they would correctly choose *mipp* when they saw an apple and *geck* when they saw a horse. Lastly, if these links generalized to the broader categories of fruits and animals, then participants would be able to label pictures of fruits and animals with words that directly co-occurred or shared co-occurrence with words *apple* (category of fruits) and *horse* (category of animals). The ability to generalize the novel words to the entire category in which familiar words *apple* or *horse* are classified would indicate a strong integration of the novel words into the semantic network.

This research will therefore provide an in-depth investigation of how mere exposure to simple co-occurrence regularities can powerfully build semantic networks that are crucial for our ability to use language.

Methods

Participants

20 4-year-old children (10 females, 10 males) from the Columbus area volunteered for this study. They were compensated with a $10 Target gift card, stickers, and an award certificate. Consent from a parent or guardian was obtained prior to the beginning of the study. The children were neurotypical and had either normal or corrected to normal vision. 21 adults also volunteered for this study. These adults were undergraduate students at The Ohio State University and were given course credit for their participation. Consent was obtained from the adult participants prior to the start of the study. Adults were neurotypical and had either normal or corrected to normal vision. One 4-year-old participant and zero adult participants were excluded from the analysis. Exclusion criteria in this case was achieving significantly below-chance accuracy in our measure of learning (see Results below).

Apparatus

The main apparatus was a display laptop on which participants watched videos, heard words and sentences, and saw pictures. All stimuli were presented using PsychoPy experimental design software (Peirce, Gray, Simpson, MacAskill, Höchenberger, Sogo, Kastman, & Lindeløv, 2019).

Training Stimuli

The primary training stimuli were sentences that were based on the two triads of words: *foobly-apple-mipp* and *dodish-horse-geck*. Each sentence used a pseudoadjective (*foobly* or

dodish) before a familiar noun (*apple* or *horse*) or a pseudonoun (*mipp* or *geck*). See Figure 2 for details. Each possible word pair (*foobly-apple*, *foobly-mipp*, *dodish-horse*, *dodish-geck*) was embedded into 10 unique sentences, for a total of 40 sentences. To ensure that any semantic links detected during the testing phase could indeed be attributed solely to regularities presented during the training phase, sentences were constructed to convey no additional information on the meaning of the pseudowords. These stimuli sentences were presented to participants accompanied by colorful, plot-free videos to keep participants' attention on the sentences.

Design

The outcome measure, or dependent variable, of this experiment was participants' accuracy during the label extension task. There were two within-subjects factors whose effects were analyzed.

Co-occurrence condition. The two levels of the co-occurrence factor are as follows: direct co-occurrence and shared co-occurrence. On some trials of the label extension task, we asked participants to label an image using either *foobly* or *dodish* to test learning of the **direct co-occurrence** link. On other trials, we asked participants to label an image using either *mipp* or *geck* to test learning of the **shared co-occurrence** link.

Learning condition. There were two levels of this factor. On some trials of the label extension task, we assessed how participants labeled images of *apples* and *horses*, which were the familiar words used in the training phase. This was done to test the **formation** of the co-occurrence links between novel words and the two familiar words. On other trials, we assessed how participants labeled images of things that were in the same category as either apples or horses, like other fruits and mammals, to test whether co-occurrence links between novel words

and the two familiar words could be **generalized** to the broader categories into which apples and horses are grouped.

Outcome. The outcome measure in this experiment was participants' accuracy during the label extension task in the testing phase. The labels that participants chose for each pictorial depiction during the testing phase revealed whether or not direct co-occurrence links and/or shared co-occurrence links had been learned during the training phase, and whether or not this learning became generalized to the broad category of the familiar nouns presented.

Procedure

Participants were tested individually in a quiet room in the lab.

Training. At the start of the training phase, participants were introduced to a fictional character named 'Jimmy' who went on a magical vacation to Zimziland. They were also told that Jimmy would tell us a story about his adventures, and use some 'silly' (novel) words along the way. Participants were then familiarized with the novel words (*geck*, *mipp*, *foobly*, *dodish*) by hearing them and repeating each one for the tester. Participants completed three blocks of training. In each training block, participants saw a video while they heard 40 pre-recorded sentences. In half of these sentences, the novel word *foobly* was paired with either *apple* or *mipp*, and in the other half, the novel word *dodish* was paired with either *horse* or *geck*. A free association task, where participants were instructed to finish 10 sentences with a word that Jimmy would use, was played directly after each video. In total, the training phase contained three videos and three free association tasks.

Testing. During the testing phase, participants completed a label extension task. Half of the trials tested whether participants linked the novel words to the specific familiar words (*apple* and *horse*) with which they directly co-occurred or shared co-occurrence. In these trials,

participants were asked to use the novel words to label pictures of apples and horses. The other half of these trials tested whether participants were able to generalize the learning of these links to other objects in the same semantic categories as apples and horses, such as fruits and large mammals. In these trials, participants were asked to use the novel words to label pictures of objects similar to apples and horses, such as a picture of grapes (same semantic category as *apple*) or a cow (same semantic category as *horse*).

Figure 3. Four trial types in the label extension task. On the right, two trials that tested formation of the direct and shared co-occurrence link. On the left, two trials that tested generalization of the direct and shared co-occurrence link.

For example, in a trial that assessed whether participants linked the novel word *foobly* with the familiar word *apple* (direct co-occurrence), we asked participants to label a picture of an apple with either *foobly* or *dodish*. A participant who chooses *foobly* to label the apple instead of *dodish* shows learning via direct co-occurrence in sentences during the training phase. In a trial that assessed the learning of shared co-occurrence links from the training sentences, we asked participants to label the same picture of an apple with either *mipp* or *geck*. A participant who chooses *mipp* to label the apple instead of *geck* shows learning via shared co-occurrence links.

On trials that assessed whether links had become generalized to encompass the semantic category to which the familiar word *apple* belongs, we asked participants to label a picture of a fruit, such as grapes, with either *foobly* or *dodish*. We also asked participants to label the picture of grapes with either *mipp* or *geck*. If participants correctly label grapes with the novel word *foobly*, we can see that the direct co-occurrence link (between *foobly* and *apple*) can actually be generalized to the broader category of fruits, and if they label the grapes with the novel word *mipp*, we can ascertain that the shared co-occurrence link (between *mipp* and *apple*) can be generalized to the broader category of fruits. See Figure 3 for details. To register participants' label extension answers, the tester used the left and right keys on a keyboard connected to the laptop display.

Results

Overview. In this experiment, 4-year-old children and adults first heard sentences in which novel and familiar words directly co-occurred or shared co-occurrence. Then, their sensitivity to these statistical regularities between words was measured using a label extension task. Our hypothesis was that children and adults were not equally sensitive to direct co-occurrence and shared co-occurrence regularities in language. More specifically, we predicted that children would be able to capitalize primarily on direct co-occurrence regularities, while adults would be able to capitalize on both direct and shared co-occurrence regularities to drive the addition of new words into the semantic network.

Participants whose accuracy in the label extension task was significantly below chance (i.e., fewer than 20 out of 48 label extension trials correct) were excluded, because below-chance accuracy is not interpretable. Ultimately, one 4-year-old participant and zero adult participants

were excluded due to failure to meet this criteria, and so analyses were calculated using 19 4-year-old participants and all 21 adult participants.

Figure 4 depicts accuracy in the label extension task. The first aim of our analysis was to compare the label extension mean accuracies to the results that would occur simply by chance (0.5). To do this, we focused on whether each age group formed the links between novel and familiar words by comparing accuracy on formation trials (direct-formation & shared-formation) to chance, or 50% accuracy. We then focused on whether participants in each group generalized these links by comparing accuracy on the generalization trials (direct-generalization & shared-generalization) to chance, or 50% accuracy. The second aim of our analysis was to discover whether label extension accuracy varied across age and/or condition. We did this by testing the effects of age (children versus adults), co-occurrence condition (direct vs. shared), and learning condition (formation vs. generalization) on mean accuracy.

Formation and generalization. We compared the following mean accuracies on the label extension task to chance (0.5) for both four-year-olds and adults: direct-formation trials, shared-formation trials, direct-generalization trials, and shared-generalization trials. Thus, eight results will be reported here. All tests were Holm-adjusted to correct for multiple comparisons.

Both children and adults demonstrated above-chance formation of direct co-occurrence links from mean accuracy values in the label extension task (child: $t(18) = 2.76$, $p = 0.039$; adult: $t(20) = 8.83$, $p < 0.001$). In shared-formation trials, only adults demonstrated above-chance accuracy (child: $t(18) = 1.19$, $p = 0.249$; adult: $t(20) = 5.14$, $p < 0.001$).

Both children and adults demonstrated above-chance generalization from direct co-occurrence links in the label extension task (child: $t(18) = 3.78$, $p = 0.006$; adult: $t(20) = 7.05$, p

< 0.001). In shared-generalization trials, only adults demonstrated above-chance accuracy (child: $t(18) = 1.90$, $p = 0.148$; adult: $t(20) = 4.18$, $p = 0.003$).

Comparing variance in accuracy across age or condition. In this analysis, we explored if accuracy varied across age groups (children vs. adults), co-occurrence condition (direct vs. shared), or learning condition (formation vs. generalization). To do this, we conducted a type II mixed-subjects ANOVA test with three factors, each with two levels (2x2x2). By running this test, we found significant main effects of age and co-occurrence condition on accuracy (age: $F(1,32) = 32.20$, $p < 0.001$; co-occurrence condition: $F(1,32) = 8.02$, $p = 0.005$). This analysis revealed that age played a significant role in accuracy, where adults were more accurate in their overall label extension performance than children. Co-occurrence condition also had a significant effect on accuracy, where accuracy on direct co-occurrence trials was overall higher than accuracy on shared co-occurrence trials. No other main effects or interactions were significant; in other words, all other p-values were above $p = 0.609$.

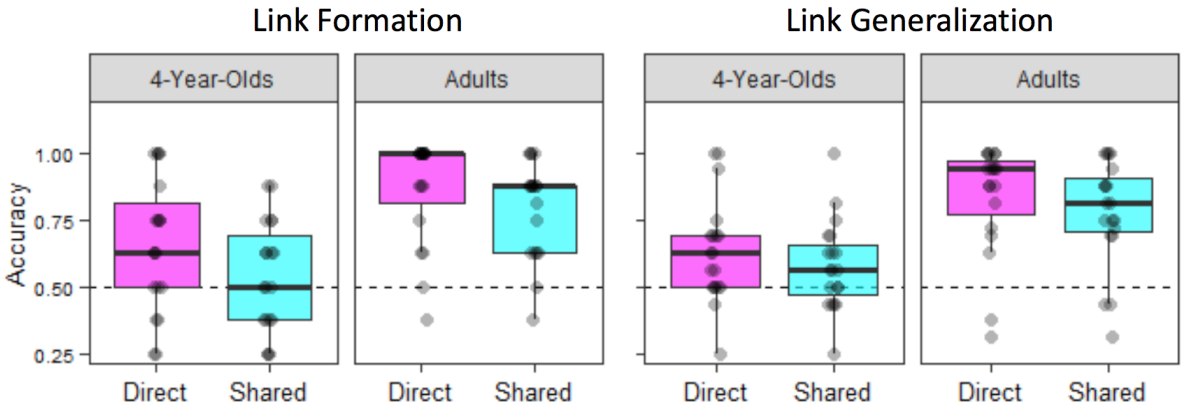

Figure 4. Distribution of accuracies for 4-year-olds and adults. In pink, accuracy on direct co-occurrence trials. In blue, accuracy on shared co-occurrence trials. On the left, direct and shared co-occurrence trials for both groups that assessed link formation. On the right, trials assessing link generalization. Each point depicts mean accuracy for one participant in a specific condition.

General Discussion

Our research was motivated by two questions: (1) can children and adults use co-occurrence regularities in language to drive the addition of new words into their lexico-semantic networks? and (2) are children and adults equally sensitive to direct and shared co-occurrence regularities in fostering the incorporation of new words into semantic networks? To investigate these questions, participants were introduced to sentences in which novel words either directly co-occurred or shared co-occurrence with familiar words. Then, we tested participants' sensitivity to these regularities using a label extension task.

A schematic depiction of the overall pattern of results is shown in Figure 5. The results revealed evidence that both children and adults linked novel words with familiar words based on direct co-occurrence. They also generalized these links to the broader range of objects in the same semantic category as the familiar word (fruit category for *apple* and large mammal category for *horse*). In contrast, only adults, and not children, formed links based on shared co-occurrence. Thus, only adults could generalize these links to the broader semantic categories to which apples and horses belong.

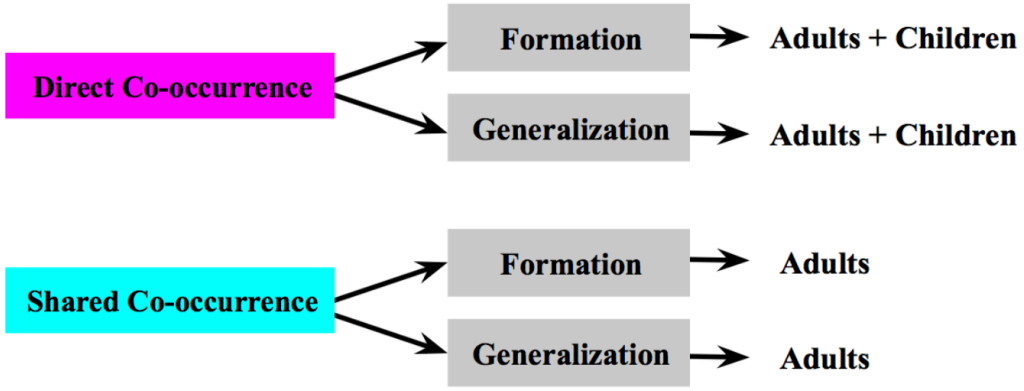

Figure 5. Schematic depicting both factors and levels for each type of trial: direct-formation trials, shared-formation trials, direct-generalization trials, and shared-generalization trials. On the right, the group(s) that were able to complete each type of trial with accuracy above chance.

These results are key to our understanding of the early sensitivity to direct and shared co-occurrence, and how this sensitivity might change throughout development. We saw that children and adults can learn from direct co-occurrence statistical regularities in language, and can also generalize this knowledge to the larger meaningful category to which familiar words belong. It is important to note that although children did form links from direct co-occurrence regularities, adults were still much more successful in both direct co-occurrence formation and generalization trials than were children. In other words, direct co-occurrence sensitivity is present in childhood but still improves with age. We also found striking developmental differences in sensitivity to shared co-occurrence. Results showed that only adults can extract meaning about novel words based on their shared co-occurrence with familiar words, and generalize this knowledge to the larger category to which our familiar words belong. However, even in adults, abilities to form links based on direct co-occurrence were stronger than the formation of links based on shared co-occurrence. Together, the results suggest that the ability to

incorporate new words into semantic networks based on direct co-occurrence develops more rapidly and remains stronger than the ability to incorporate new words based on shared co-occurrence.

Contributions of Co-Occurrence Regularities

As reviewed in the Introduction, previous studies on direct and shared co-occurrence in a variety of domains like language, sounds, and visual cues paint a picture of sensitivity to direct co-occurrence regularity as an ability that emerges early in development, and sensitivity to shared co-occurrence regularity as a cognitive skill that may not be present at the onset of development, but evolves as one matures (Saffran, Aslin, & Newport, 1996; Fiser & Aslin, 2002; Wojcik & Saffran, 2015; Saffran and Benitez, 2018; Bauer & San Souci, 2010; Bauer & Larkina, 2017; Schlichting et al., 2017). Our analysis further elucidates the relationship between direct and shared co-occurrence regularities, as well as the relationship between co-occurrence and novel word learning. Prior studies portrayed sensitivity to direct co-occurrence as a phenomenon that can be captured as early as infancy (Saffran, Aslin, & Newport, 1996; Fiser & Aslin, 2002). Other researchers found that direct co-occurrence sensitivity in toddlers and children ages 4-8 was even stronger, and postulated that this sensitivity may be able to support word learning (Wojcik & Saffran, 2015; Saffran and Benitez, 2018). The present results are consistent with the previous studies, and expand on them further by showing that four-year-olds could learn some meaning of a novel, unfamiliar word through its direct co-occurrence with a word that they knew. This finding has far-reaching implications regarding the mechanisms present during this early stage of development for expanding the semantic network.

Prior evidence concerning sensitivity to shared co-occurrence regularities has suggested that it may be weak or absent in four-year-olds, but that it is a skill that undergoes development.

For example, between the ages of 4 and 6, researchers found a marked improvement in the ability to integrate facts across separate episodes (Bauer & San Souci, 2010). By age 8, this skill had improved yet again (Bauer & Larkina, 2017). By adulthood, sensitivity to shared co-occurrence between words and in other domains appears robust (Schlichting et al., 2017; McNeill, 1963; McNeill 1966). Our findings corroborated these prior findings, as only adults, and not children, were sensitive to shared co-occurrence regularities presented in our stimuli. Moreover, adults showed that they could capitalize on shared co-occurrence regularities in language to build meaningful links between familiar words and new words, and could even generalize this knowledge to objects not presented at training (*apple* and *horse*). The present findings are impressive, as they show that the ability to form semantic links from shared co-occurrence regularities improves over time and can support generalization of knowledge.

Limitations

One limitation of the present study is that the label extension task in the testing phase required participants to explicitly express their choice of label for each trial. As is the case with an explicit label measure such as this one, it may have taken a more concerted effort on the part of the participants, especially four-year-olds, to verbalize their response. If making these responses was harder for children than adults, this could have increased their number of errors, and therefore underestimated their formation and generalization of links based on co-occurrence. Thus, the label extension task may not capture all learning that occurred.

Another limitation comes from the fact that this experiment used four-year-olds and university undergraduate students: two groups whose ages were highly different. It can be assumed that a diverse array of cognitive skills surface and improve between four years of age and adulthood. Given the large age differential between the two groups and the scope of our

experiment, there is no clear answer as to exactly when the ability to form and generalize links from shared co-occurrence regularities maturates.

Further Directions

As stated previously, an explicit labeling measure may pose additional challenges to participants and therefore may underestimate learning, especially in children. More sophisticated measuring tools may provide a clearer understanding of the learning taking place. Testing learning with a more sensitive, millisecond-by-millisecond measure, like eye-tracking, allows for more accurate data acquisition. Eye-tracking also unobtrusively and implicitly measures underlying cognitive processes, like semantic activation of specific objects or words if a linked word has been heard, with no verbalization of answer choices necessary from participants. A future study using eye-tracking might replicate this study's training phase, and replace the label extension task with an eye-tracking measure. Using the eye-tracking paradigm, participants would hear the novel words they have learned during training (*foobly*, *dodish*, *mipp*, *geck*) and see images of the familiar words *apple* and *horse*. During this task, their eye-movements would be recorded. If a new link had been formed between the novel word and the familiar word with which it directly co-occurred or shared co-occurrence, the participant would most likely look at the corresponding picture on the screen. Thus, eye tracking could measure links without requiring explicit responses.

In addition, the age difference between our two groups of comparison leaves the question unanswered of when more complex co-occurrence sensitivity and generalization ability unfolds. Future experiments could answer this question with a similar methodology to ours, but with differently defined age groups. From our study, we know that by age 4, the ability to learn and generalize links from direct co-occurrence regularities in language have begun to emerge.

Furthermore, we know from previous studies that sensitivity to statistical regularities can change between the ages of 4 and 6, and between the ages of 6 and 8 (Bauer & San Souci, 2010; Bauer & Larkina, 2017). Therefore, a study that used the methodology described in this experiment with an experimental group of four-year-olds, plus groups of six-year-olds, eight-year-olds, and ten-year-olds (or even older) may provide more insight into the developmental trajectory of abilities to form links between words based on direct and shared co-occurrence.

Lastly, we treated all generalization trial images as related to familiar words *apple* and *horse* to the same extent. For example, we treated an image of a zebra with the same likeness to a horse as we treated an image of a cow. A new experiment could be proposed which segments 'related' images in the same semantic categories as the familiar words based on their physical similarities to the familiar words. Using this design, a stimuli image of a zebra ('near') would be treated differently in the analysis than would an image of a cow ('far'), since a zebra is more similar to a horse than is a cow. Considering these near and far generalization conditions would allow researchers to test whether the degree of similarity of the image (zebra = more similar, cow = less similar) to the familiar word affects how participants generalize these links.

Conclusion

The aim of this research was to uncover whether new words could be added into semantic networks based on statistical co-occurrence regularities in language. We found evidence that sensitivity to statistical direct co-occurrence regularities in language was present early in development, by 4 years of age. Sensitivity to shared co-occurrence regularities in language, however, is protracted; it was only present in adults. Importantly, when people were sensitive to co-occurrence regularities, we found that novel words became linked not just to specific familiar

words (apple and horse), but also to objects in the broader semantic network of the familiar words. This research provides an in-depth investigation of how mere exposure to simple co-occurrence regularities can powerfully build semantic networks that are crucial for our ability to use language. Our study shows that these regularities can support new word acquisition in children and adults, and that abilities to link new words into semantic networks based on co-occurrence regularities improve over development.

References

Bauer, P.J., & Larkina, M. (2017). Realizing relevance: The influence of domain-specific information on generation of new knowledge through integration in 4-to 8-year-old children. *Child Development*, *88*, 247-262.

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology*, *107*, 452-465.

Benitez, V., & Saffran, J. R. (2018). Predictable Events Enhance Word Learning in Toddlers. *Current Biology*, *28(17)*, 2787-2793.e4. https://doi.org/10.1016/j.cub.2018.06.017.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822-15826.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1-37.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211.

McNeill, D. (1963). The origin association within the same grammatical class. *Journal of Verbal Learning and Verbal Behaviour 2*.

McNeill, D. (1966). A study of word association. *Journal of Memory and Language*, *5*, 548-557.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Savic, O., Unger, L., & Sloutsky, V. (2019). Rapid semantic integration of novel words following exposure to distributional regularities, *Proceedings of the 41th Annual Conference of the Cognitive Science Society*.

Savic, O., Unger, L., & Sloutsky, V. Becoming Organized: How Simple Learning Mechanisms may Shape the Development of Rich Semantic Knowledge. PsyArXiv [Preprint]. [cited 2020 November 11]. Available from psyarxiv.com/5tpf2.

Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*, *29*, 37-51.

Sloutsky, V. M., Yim, H., Yao, X., & Dennis, S. (2017). An associative account of the development of word learning. *Cognitive Psychology*, *97*, 1-30.

Wojcik, E. H., & Saffran, J. R. (2015). Toddlers encode similarities among novel words from meaningful sentences. *Cognition*, *138*, 10-20.